Final Project Report: Optimizing Course Arrangement for Faculty Preferences

**University of Southern California**

**MS Business Analytics**

**DSO 570: The Analytics Edge: Data, Models, and Effective Decisions**

**Spring 2020**

# **Executive Summary**

Scheduling classes for the USC Marshall School of Business is a complex process that involves balancing the preferences of students, faculty, and departments, while also keeping in mind factors like class sizes, building availability, and special requests. Currently, department course times and rooms are allocated starting almost a year prior, after which departments themselves fill in their respective slots in Phase I of the process. Adjustments are made in Phase II, but there is currently no structured system in place to systematically prioritize and accommodate these concerns. Therefore, our team has discovered the following opportunity for improvements and created an effective course scheduling system that aims to streamline the process further and achieve the following goals:

* Reduce the human effort in USC Marshall course allocation and scheduling that is currently being handled by course coordinators and individual departments to ease their workload
* Prioritize the preferences of faculty specifically to create a course schedule that satisfies these constraints and can effectively serve as a tool with which to schedule classes for the future

Since there is abundant data available for Marshall faculty preferences, we decided to use faculty preferences as the main priority when creating our scheduling system, while also trying to accommodate other preferences (student, department, etc.) as much as possible**.**

Since significant improvement can be gained by satisfying such preferences has been observed, we provide the following recommendations:

* Systematically organize all possible preferences over time slots for all courses and professors. There are various soft requirements of preferences not considered in the current model but can be incorporated in the further models. Shannon and Hal’s team and department coordinators can expand the optimization model with the organized preferences in numeric values and set leverages according to practical experience.
* Sort out all kinds of durations for various kinds of courses. Various kinds of course units and durations do not quite follow regular rules, so Shannon and Hal’s team and department coordinators can organize the actual number of slots for each course, in which slots should be a smaller unit like half an hour instead of 2 hours, and incorporate new time slot divisions in the original optimization tools.

# **Opportunity for Improvement**

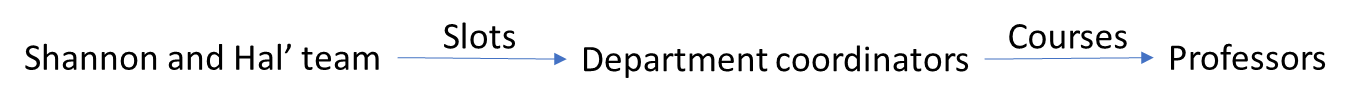
1. **Background Preparation**

Before discussions on opportunities to improve the scheduling of courses and classroom allocations at USC Marshall, we would like to elaborate some of the more important points in the current scheduling process that will lay the background to how advanced analytics can be used to advance the system, benefitting Shannon and Hal’s team, senior administrators, and potentially all other parties concerned.

1. **Summary of Scheduling Process**

The scheduling system of courses and classrooms at USC Marshall involves three parties: Shannon and Hal’s team, department coordinators, and professors. The overall procedure is as follows:

Shannon and Hal’s team will try to distribute all available classrooms and time slots to departments according to the experience and schedules of previous semesters and consideration of possible numbers and types of courses from a total of 7 departments including Business Communication (BUCO), Data Sciences and Operations (DSO), Finance and Business Economics (FBE), Leventhal School of Accounting (ACCT), Management and Organization (MOR), Marketing (MKT), and Lloyd Greif Center for Entrepreneurial Studies (BAEP). Besides, Shannon and Hal’s team will assign classrooms and time slots one year ahead for the same semester as the current semester. After initial distributions of classrooms and time slots without courses on the department level, the assigned empty slots are handed over to department coordinators. Department coordinators populate the allocated time slots with departments’ educational plans of courses. Meanwhile, coordinators will negotiate with professors to assign proper course plans to professors and hear the opinions and preferences from professors about courses and time schedules. In conclusion, the cooperation between Shannon and Hal’s team, department coordinators, and professors goes into linear directions of assignments and connections as the following map illustrates:



In addition, in the traditional process, the steps described above are marked as Phase 1 to populate most of the courses to time schedule and classrooms. To deal with additional courses and time slots not assigned, Shannon’s team works together with department coordinators to schedule the remaining courses. This phase is called Phase 2, but as part of the opportunities of improvements, we will try to mitigate the influence of second-course distributions with the automated optimization tools and improve the efficiency by directly assigning courses to time slots and classroom, which is equal to merging Phase 1 and Phase 2, and quickly updating course schedules with quick updates of courses and preferences and optimization.

1. **Project Goals**

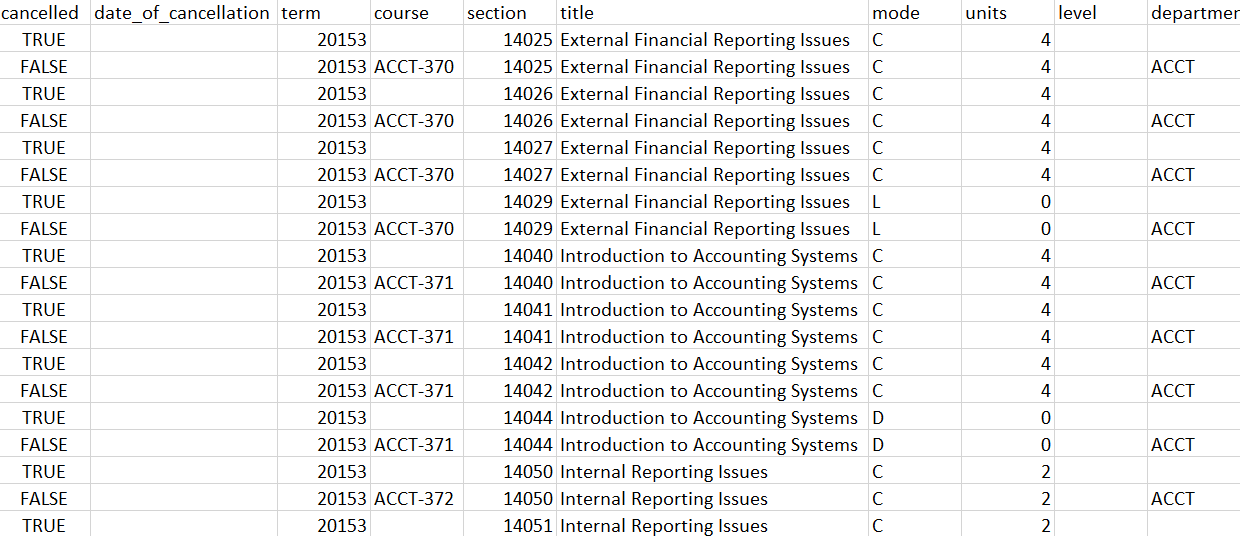
Nowadays, the scheduling of courses and classrooms at USC Marshall approaches issues, such as the rising number of courses. Shannon and Hal’s team and department coordinators would like to accept advanced business analytics techniques and faster optimization methods to deal with the increasing complexity of the data available, constraints, and course changes. We will focus on how our optimization tools will solve the following questions:

1. Efficiently utilize all time slots and classrooms by providing a final schedule covering courses’ allocations to specific time slots and classrooms. This remains our priority as our main constraints are the limited number of classrooms available, the increasing number of courses, and the strict time restrictions for the middle of one day. In the end, the solution will reduce the current scheduling team’s workload spent on referring to previous schedules, manually assigning time slots, and negotiating with professors.
2. Satisfy the preferences of professors. It is important for a better environment of teaching, which is good both for professors and students. One may regard professors’ preferences as negotiable and consider them soft constraints, but we take them in as a high priority because, currently, no systematic way of accommodating faculty preferences exist.
3. Quickly deal with sudden, new changes in terms of professor or course schedules. Even after Phase 1 (or possibly even 2), course demands, assigned faculty, classroom availabilities, and such can suddenly change. Such changes required a lot of manual labor from current administrators. Thus, our solution will provide a way for such issues to be dealt with quickly and efficiently.
4. Reduce the workload of Shannon’s team and department coordinators. They will spend much of the efforts in referring to the previous schedules, assigning time slots, and negotiating with professors, so an automatic optimization tool certainly will save the efforts of Shannon’s team and department coordinators and improve the efficiency of scheduling and reflecting the preferences.
5. **Discovery of Inefficiency**

Because the scheduling system of courses and classrooms at USC Marshall is based on current administrator experiences and the schedules of previous semesters, a significant lack of efficiency exists. The current two-phase process of populating most courses and then trying to fill in remaining courses as well as the complex negotiation required between the three parties of Hal and Shannon’s team, department coordinators, and professors (not to mention student demand) all further complicate course scheduling, laying the potential for further inefficiency. By exploring previous course schedules and the records of how they were done, we discover several specific points of inefficiency:

1. **Heavy Workload and Manual Inefficiency**

We focus on not only the courses that were decided on previous course schedules but also courses that appeared and disappeared in the data entry records because the data entry records well reflect the whole process of assigning and changing courses according to emerging background changes and how many rows of course scheduling records were edited, which reflects workload in manual editing of adding or canceling courses. Our data entry records span from 2015 to 2019, consisting of Spring, Summer, and Autumn, with a total of 7441 data entry records. An illustration of part of the data entry records is as follows:



Here we are concerned about the column “cancelled”, which represents whether the section was later canceled. If a course is canceled, probable reasons for cancellation include data entry error and insufficient demands. We notice that a noticeable portion of rows contains canceled labels, which questions the efficiency of data entry and course scheduling.

We further explored the total number and proportion of canceled course records, as well as their distributions in all semesters. A summary table of the total canceled records and the percentage in all data entry records is illustrated as follows:

|  |  |
| --- | --- |
| Total Cancelled Records | Percentage |
| 2534 | 34.05456255879586% |

We discover that the total canceled records are pretty significant, with 2534 canceled records amounting to around 34% of all data entry records. This signifies the problem of large-scale modification of course schedules, either due to data entry errors or insufficient demands. Such frequent modification of course schedules definitely means inefficiency due to human errors and manual editing of adding and canceling courses.

Meanwhile, we do not ignore the total course records and total canceled records for each semester. The number of course records, the number of canceled records, and the percentage are as follows:

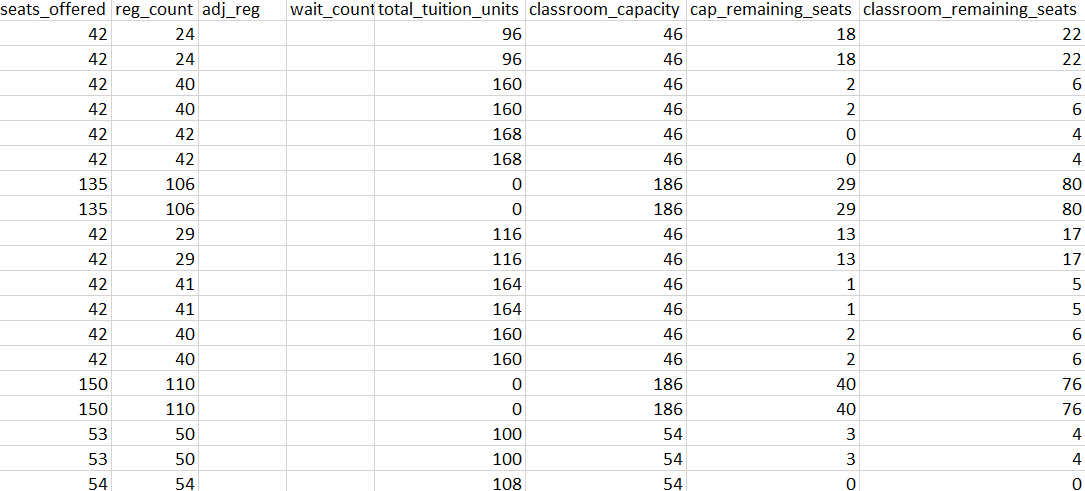


We discovered that the average number of courses that are successfully scheduled for practice is around 600, which itself is a serious workload to Shannon’s team and department coordinators for each semester. Meanwhile, we notice that the phenomenon of modification over canceled courses is significant from Fall 2015 to Spring 2017, where nearly all courses were canceled and rearranged. Whether it was due to data entry errors or large and frequent adjustments of the course schedule, it was a heavy workload for Shannon’s team and department coordinators. Even though such large-scale modification has not happened since then, we are still alerted about the possibility of large-scale data entry errors or modification of course schedules from the demands of students and professors.

In conclusion, due to over 600 courses to be scheduled and the possibility of large-scale modifications and errors illustrated above, the heavy workload becomes a critical factor of inefficiency, which brings forth the needs of automatic optimization.

1. **Low Utility of Classroom Capacity**

It is claimed that there are no additional classrooms due to building restrictions, and it is hard to schedule all courses with suitable time slots. Without outside resources, we should look into whether the previous course schedules utilized classroom seats as much as possible. We imagine that it is difficult for parties involved to investigate whether courses have been optimally utilizing most classroom seats. This is due to the continued usage of previous course schedules, fast-evolving course plans from departments and programs, and frequent changes of course schedules. An illustration of data points related to classroom seats and course seats is as follows:



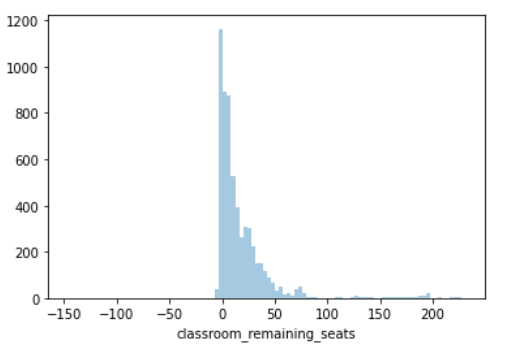
We observe that a considerable number of classroom seats are not utilized by a semester’s courses, thus leading to a lower utility per classroom seat used.

A deeper exploration into the total unutilized classrooms seats and the percentage across total seats for each semester can be seen below:



We discover from the percentages of unutilized classroom seats that around 30% of the total classroom seats are empty for previous course schedules, which means some courses do not match the sizes of classrooms, and many large classrooms are forced to incorporate courses with smaller registration seats.

Meanwhile, we further explore the distribution of unutilized classroom seats for all courses, which is illustrated as follows:



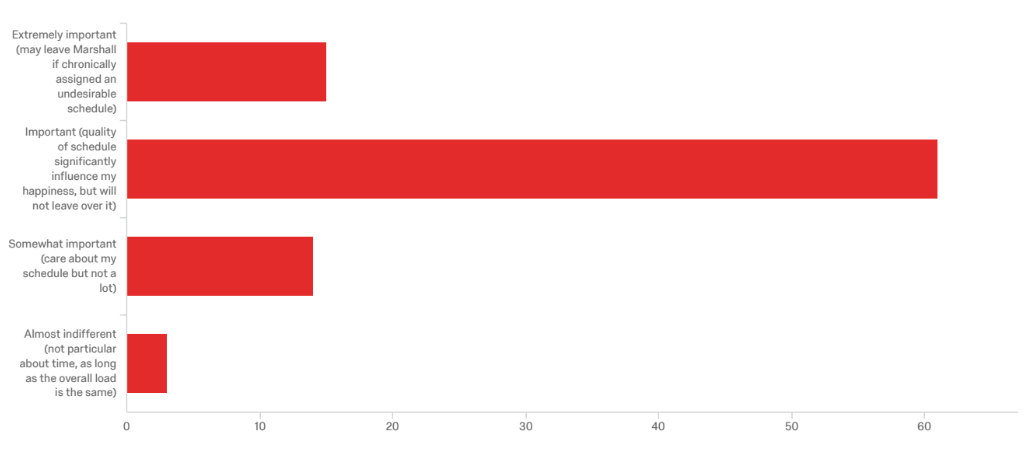
We notice that a significant portion of empty classroom seats exceeds 20, and there are cases where the registered seats exceed classroom limits, or the classroom is too large to be utilized by courses.

In conclusion, because of the large space that is not utilized by course schedules, the inefficiency arises from the limited use of classroom space, which is what our optimization tool will focus on.

1. **Lack of Systematic Preference Consideration**

Due to the complicated negotiation between Shannon’s team, department, coordinators, and professors, the professors can express their strong preferences over course selection and time schedule, but the process will continue for a long time. It is admitted that currently, there is not a systematic method of accommodating faculty preferences.

Moreover, dissatisfaction would bring potential harm to professor positions. According to the faculty survey, some of the professors would hold extreme attitudes toward course schedules. An illustration of the professors’ attitude towards course schedules is as follows:



We observe that around 60% of professors would be affected by schedules, and over 10% of the professors could leave Marshall if the schedule is undesirable, which is a harm to the continued teaching and research of Marshall School of Business.

In conclusion, there is an urgent need to set up an optimization tool that will systematically incorporate faculty preferences.

1. **Conclusion and Approach**

Based on the data analysis above, we conclude 3 causes of inefficiency in the process of course scheduling for USC Marshall:

* The heavy workload of manual editing courses and the possibility of large-scale errors
* The low utility of classroom capacity of 30% empty seats and excess of capacities
* Lack of systematic arrangement towards the preference of professors

Our optimization tool aims to solve the three problems with the three corresponding approaches:

* Utilize the automatic power of Python and Gurobi to optimize the course schedule quickly based on fast changes of courses and preferences
* Set optimization of classroom utility as the goal of the objective value
* Incorporate preferences of professors in the objective value as a systematic and mathematical way of optimizing the needs of professors

# **3. Optimization Methodology**

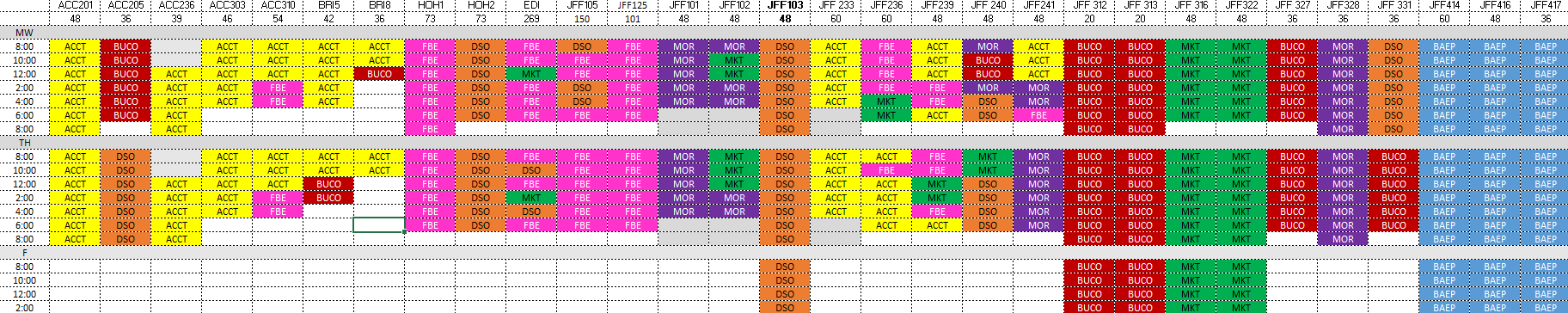
Based on the opportunities for improvements in fully utilizing classroom seats and stabilizing classroom schedule under adding and canceling courses, we strongly suggest replacing or improving the traditional methods of adding courses on available time and negotiating with coordinate separately with integrated optimization methods, which is faster than negotiations, easy to discover potential solutions and directly reflect needs of professors over automatic scheduling.

We describe the demand from professors and course schedules as quantitative problems as part of the optimization process, and we utilize the power of Python and the optimization tool Gurobi to realize the demand, the goal, and the constraints and automatically reach maximum solutions.

1. **Input Data for Optimization**

Our optimization method takes information from four aspects: courses, time schedules, classroom information, and professor preferences over courses. The logic is that we must arrange the courses to a time schedule and empty classrooms under the preference of professors.

1. Time schedule: we must know what time is available for course arrangements. According to the common schedule of Marshall School of Business for courses, courses are available for such time slots: from 8 am to 9:30 pm, a total of 7 slots of nearly 2 hours for from Monday to Thursday, and 8 am to 2 pm for Friday, a total of 4 slots. We arrange the courses based on this available schedule, and an illustration of a standard time schedule is as follows:



Those slots of 2 hours are for further arrangement of coordinators. In another version of course arrangement, slots follow a duration of 1.5 hours, so there are 9 slots from Monday to Thursday, and 5 slots for Friday. Course duration is decided by course credit, approximately 1 credit for half an hour. Usually, occupying two same slots as Monday-Wednesday or Tuesday-Thursday for such courses is also a good option because it reduces the trouble of moving around across classrooms or days for students. We put the suggestions into the considerations of our objective.

1. Course: we must get the list of courses that would be put into our schedules. These courses would contain the following information for further optimization: credits and capacity. The credits measure how many slots would be occupied and capacity measures which classroom would fit the course needs and how to utilize the classroom seats. The courses should match the preference of professors to maximize the preferences of professors.
2. Classroom information: We must get the list of classrooms available for course arrangements. As mentioned above, classrooms restrict seats and measure seat utility to hold more courses. The capacity of classrooms is enough for input data.
3. Preference of professors: The preferences of professors are closely connected with the arrangement of courses, and the needs of professors are various, including time schedule limits and position-related dissatisfaction over unsuitable schedules. We need the following concerns from the professors about the courses:
   1. Whether courses are after 10 am and before 6 pm
   2. Whether the courses prefer Monday and Wednesday,
   3. Whether the courses prefer Tuesday and Thursday

Each of the seven preferences is measured by numeric values. Following the survey for faculty, we rank the preferences using 4 integers from 1 to 4. Those preferences match corresponding courses the professor oversees, and they place leverage over our objective goals.

1. **Output of Optimization**

Our target is to optimize the arrangement of courses to the classrooms and their available time slots, so we will illustrate how courses are placed in classrooms and time slots. The output will be a table with rows as time slots and columns as classrooms, showing course names in the table, which pretty much resembles the illustration in the time schedule part above.

1. **Decision Variables, Objectives, and Constraints**

Our decision variables, depending on the selections of classrooms and time slots, are whether a certain course is placed in the specified time slot and the specified classroom , where the course belongs to the set of courses , the classroom belongs to the set of classrooms , the time slot belongs to the set of time slots .

In addition, we need to mark the various data mentioned above. We introduce seven preferences for each professor or course mentioned in the section of preference of professors as which belongs to sets of containing numeric preferences with the ranking range of 1 to 4. In addition, the required seats for each course is marked as , belonging to the set of all course seats , the capacity for each classroom is marked as , belonging to the set of all classrooms capacities , and whether the course is well arranged with the same time slots on Monday-Wednesday or Tuesday-Thursday time slots is marked as , which belongs to the set of the same conditions for all courses as . The credits, or the time slots needed, are represented by .

Our objective goal will be a mixture of leverages over various constraints and preferences from professors and classroom utility. Because we are trying our best to satisfy all preferences, regulations, and constraints, in which some are soft requirements, we have multiple objective goals. Those goals consist of two parts, which correspond to the two improvements that are discussed above and going to be improved by optimization methodology:

* 1. Maximizing capacity utility: as mentioned above, on the one hand, 23% of the classrooms have more than 20 empty seats, which are not utilized enough. On the other hand, more courses are populated into time schedules. Thus, we will measure how we utilize classroom capacities. We measure the utility using the sum ratio of total seats used by each course and the total capacity of each classroom to measure the average utility percentage of courses.
  2. Satisfying the preferences of courses: the preferences of courses are regarded as rather soft and negotiable conditions, so compared with strict constraints like classroom capacities and course conflicts, we incorporate those preferences into the objective goal and add leverages on the preferences as soft boundaries which take effects in the optimization process. We tend to give the preferences with equal leverages, but it is also optional to adjust the leverages. Here we assume all preferences are equal, and we calculate the total benefits brought by satisfying preference, which can be represented by preference score. If we satisfy one preference, the preference score can be added to the objective value. As each course will satisfy some of the preferences, we will add up all preferences that are satisfied to represent the total gain for satisfying preferences, which we intend to make it as high as possible by optimizing it in the objective value. In addition, the preference of Monday-Wednesday or Tuesday-Thursday slots is also a soft boundary, so we assign the preference leverage as the same with the highest preference score 4 to get equal leverage with the preferences of professors.

Our objective goal is to maximize the sum of the two improvements. To leverage the importance of capacity utilization and preferences of courses, we give the capacity utility the same importance as the highest preference score 4.

Our constraints are less than expected because most of the soft requirements are converted into objective elements. In total, 5 constraints are added to our optimization process:

1. Room Capacity: the maximum number of students for the course cannot exceed the limit of the specified classroom.
2. Room Time Conflict: the time slots or the classrooms for courses should not conflict with each other. It is basic that two courses should not appear at the same time in the same classroom.
3. Total Course Time: the course should satisfy the weekly duration of enough time slots, which is decided by course credits.
4. One Course Room: only one classroom should be used for each course so that the students would not waste time moving to another classroom for the same course.
5. Binary: whether a certain course is placed in the specified time slot and the specified classroom, which should be a binary value.

A more accurate and mathematical abstract formulation will be illustrated in Technical Appendix for further discussion of details like leverage between capacity utilization and preferences of courses in more analytical ways.

1. **User Instruction**

There are two important things to note for the instruction of users:

* 1. This optimization involves complicated calculations towards the final solution of 3-dimensional information of classrooms, courses, and time slots. Thus, the optimization is built on Python and Gurobi, which is not user-friendly towards non-programming faculty and coordinators. For more information about the technical structure and functions of the optimization tools, please refer to the documentation of optimization tools in Python and Jupyter Notebook files as well as the sample inputs and outputs.
  2. This optimization is intended for the integration of the preferences of professors, which are collected by coordinators and negotiated by both Shannon and Hal’s team and coordinators, and classroom maximum utility to hold more courses, which is managed by Shannon and Hal’s team. It improves the efficiency of incorporating professors’ opinions, so it is intended available for both Shannon and Hal’s team and coordinators to better arrange the preferences and courses and make fast and accurate changes to the schedule.

The interaction with the optimization tool is conducted via Python Console intended for easy coding and use. We use the following command to start the optimization tool and process:

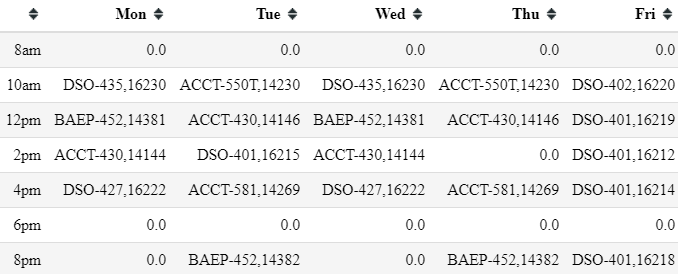
python optimize.py input.xslx output.csv

Here optimize.py is the optimization tool sealed in Python codes. Input.xlsx is the input file containing all needed data of course credits, classroom capacity, and course preferences. Output.csv is the output file containing the final schedules of courses for time slots and available classrooms.

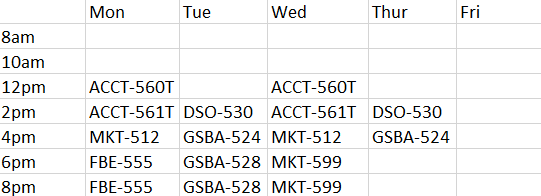
We expect that Shannon and Hal’s team and department coordinators will cooperate in assigning professors to the course tasks first and match the preferences of professors to the courses, so a joined table containing both preferences and course information is available for optimization. By updating the table with new or canceled courses and changes of preferences of professors, Shannon and Hal’s team and department coordinators can react more quickly and incorporate the needs of professors by using optimization tools, which solves the problem of no systematic way of eliciting and accommodating faculty preferences before.

# **4****. Optimization Results**

We successfully built the optimization tool in Python and Gurobi following the formulation of the course scheduling problem, and we are ready to verify our idea of improvements on simulated datasets. We used a sample dataset, sample\_input.xlsx, to investigate the details of the optimization and discover points for improvement. An illustration of the optimized schedule is shown below:



Meanwhile, in contrast, the original schedule of Fall 2016, from which the sample was taken, is shown below:



Instead of empty course arrangements in the morning and full course schedules in the evening, which professors find undesirable, the optimized schedule can hold more courses than the original by optimizing the preferences of pair slots on weekdays and time slot preferences over time and weekdays and dealing with one-unit courses with less priority.

This improvement can be quantified with multiple values. A summary table of relevant quantitative values are shown as follows:

|  |  |  |
| --- | --- | --- |
| Leverage of Utility | Leverage of Preference | Preferences Fulfilled |
| 24.363636363636378 | 109 | 34 |

We can see that while the optimized schedule maintains the utility of setting 22 slots of around 24, the optimization gains largely from satisfying the preference of courses. With 34 preferences fulfilled, an average of more than one preference is satisfied for each course. Moreover, they accurately pick up high preferences, with averages of more than 3 for each preference. Although we could not quantify the benefit of preferences for original schedules, we still see great gains from utilizing the preferences in optimized schedules.

Moreover, seeing how even more time slots are fulfilled in the optimized schedule, we learn that course scheduling is not as restricted by the number of available classrooms as may have been previously thought. There is no need to require more buildings and classrooms or improve the utility of classrooms. The main obstacles come from the time arrangement of courses, in which most of the problems come from time preferences. Although the soft boundaries are negotiable, maximizing the incorporation of preferences and convenience becomes the focus of Shannon’s team, department coordinators, and optimization tools.

# **5. Discussion**

1. **Appropriateness of Methodology:**

In this project, our goal is to arrange the courses for proper time slots and classrooms with the largest utility used under the preference of professors. So, we consider our input data from information about the courses, time slots, classroom capacity and seats, and professor preferences. This also helps resolve the inefficiency in the current scheduling process we mentioned in part two. Our output data will be a table with rows as time slots and columns as classrooms, which shows courses’ names in the table. This clearly and directly shows our result and fits the goal of our project. The constraints and objectives are also based on the goals of our project. For the constraints, we consider room capacity, room time conflict, total course time, one room rule for each course, and binary variable. The objective is to maximize the sum of two main improvements, the capacity utilization and the preferences of the courses.

The basic assumption of the optimization formulation and tool is that while considering maximizing the utility of classrooms, the preferences of courses are soft requirements and can be incorporated by maximizing the estimated gains of satisfying preferences as much as possible in the maximized objective values. The possible weaknesses are that besides the four preferences incorporated in this optimization model, more kinds of preferences should be considered, and more different time slots decided by course units, and course properties should be incorporated instead of fixed 2-hour slots.

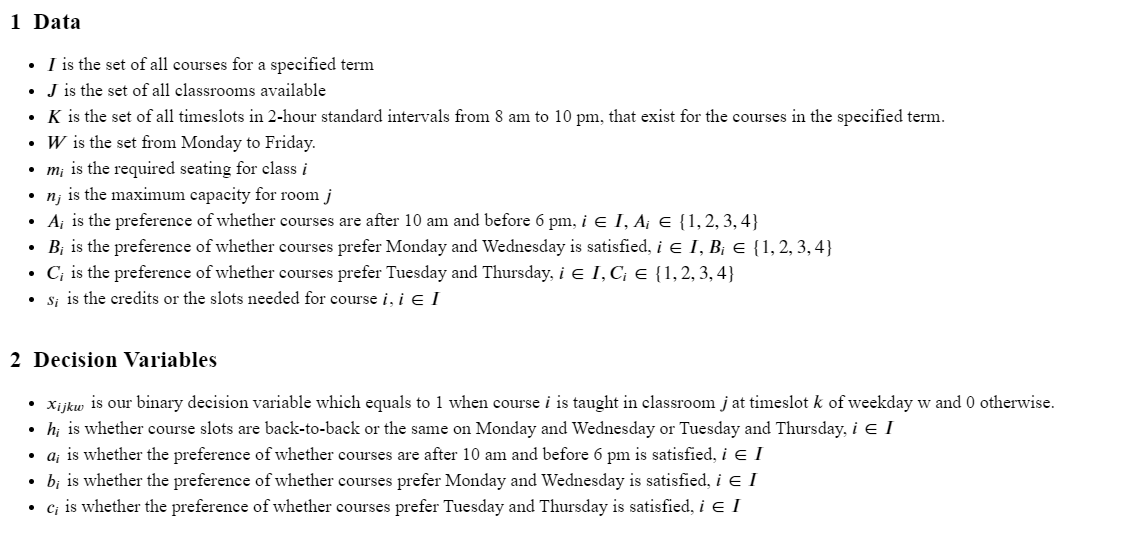
1. **Final Recommendation:**

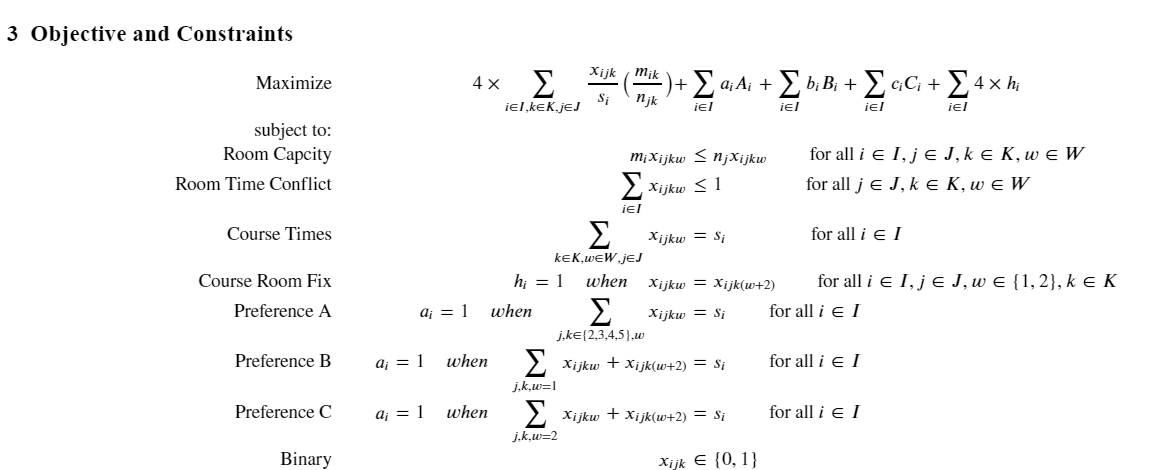
My final recommendation towards the target user, Shannon and Hal’s team and department coordinators, would be the following next steps:

* Systematically organize all possible preferences over time slots for all courses and professors. There are various soft requirements of preferences not considered in the current model but can be incorporated in the further models. Shannon and Hal’s team and department coordinators can expand the optimization model with the organized preferences in numeric values and set leverages according to practical experience.
* Sort out all kinds of durations for various kinds of courses. Various kinds of course units and durations don’t quite follow regular rules, so Shannon and Hal’s team and department coordinators can organize the actual number of slots for each course, in which slots should be a smaller unit like half an hour instead of 2 hours, and incorporate new time slot divisions in the original optimization tools.

**Technical Appendix**

**A1. Mathematical Formulation**





**A2. Discussion of Technical Details**

This part is used to discuss technical details of the assumption of the optimization model and further extends the discussion on possible improvements.

The basic assumption of the optimization formulation and tool is that while considering maximizing the utility of classrooms, the preferences of courses are soft requirements and can be incorporated by maximizing the estimated gains of satisfying preferences as much as possible in the maximized objective values. It is accepted because there are huge amounts of preferences from students and professors, which usually conflict with each other, so setting the preferences as constraints is not a viable option. Moreover, setting preferences as the optimization goals make the optimization process satisfy as many preferences as possible and solve the possible conflicts by arranging priorities with leverages to different preferences in the objective goals.

The possible weaknesses are that besides the four preferences incorporated in this optimization model, more kinds of preferences should be considered, and more different time slots decided by course units, and course properties should be incorporated instead of fixed 2-hour slots.

For more preferences to be included in the model, there are more preferences like whether the weekday is specified, whether the course should be back-to-back, whether the course is a half-semester course, and more. Especially considering some of the professors dissatisfied with bad schedules tend to leave Marshall School of Business, which is a harm to teaching and research, preferences like this should also be included and put strong leverage on them to make them satisfied by optimization. If time and resources allow, the problem can be fixed with more investigation into various needs of professors and formulate the preferences and leverages based on the list of preferences.

For more different time slots, the course units range from 0 to 15 credits, but the duration of the course is not in a linear relationship with the units. For example, 3 units usually correspond to 3 hours each week, but some of the courses like BUAD-301 occupy 4 hours each week. Courses with 0 units use 2 hours or else each week. Moreover, there are circumstances where courses are taken for half-semester or else. Considering the various durations above, a new time schedule used to populate courses in the optimization model should replace the original 5-day-7-slot timetable, and the new timetable is able to incorporate all circumstances of time schedules. If time and resources allow, the problem can be fixed by requiring the rules of converting units to the duration or directly requiring inputs of course duration. We will use a smaller slot like half an hour and more variables like required weeks to expand the timetable and incorporate more circumstances.